DOCUMENT RESUME

ED 279 002 CS 210 258

AUTHOR Brossell, Gordon
TITLE Essay Test Topic Development.
PUB DATE Apr 86
NOTE 23p.; Paper presented at the Annual Meeting of the Conference on Writing Assessment (4th, Cleveland, OH, April, 16-18, 1986).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Higher Education; *Research Methodology; *Research Needs; Secondary Education; *Test Construction; Testing; Test Items; Test Reliability; *Writing Evaluation; *Writing Research; Writing Skills

ABSTRACT

Extensive research efforts in the past decade have measured academic proficiency and progress in both theory and practice of large-scale writing assessment. The following major situational variables were identified in these investigations: (1) topic variables such as writing prompts, subject matter, mode of discourse, length, and rhetorical specification; (2) writer variables--such as topic interpretation, writer anxiety, and writing assessment episode--addressed by anthropological, sociological, linguistic, and psychological research; and (3) common procedural variables likely to affect the outcomes of writing assessments, including writing stimulus, instructions, common environment, time limit or guide, and scoring system. Further research efforts are needed to investigate the refinement of the technical procedures for operating and scoring writing tests, the interaction between writers and various writing assignments, and the relationships between the goals of writing assessment and the measures of writing performance provided by assessments. Most importantly, future investigations should examine ways to build conceptual bridges between the outcomes of assessment and the processes whereby writing is taught and learned. The key to this construction lies in determining how to create the conditions under which good writing is known or is apt to occur. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279002

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Essay Test Topic Development

Gordon Brossell

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gordon Brossell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

CS 210258

Anyone concerned with the theory and practice of large-scale writing assessment may now turn to a growing body of research knowledge that has emanated from the extensive effort to measure academic proficiency and progress during the last decade. On the assumption that it is valuable to know at what points the widespread practice of an enterprise converges with or diverges from research findings about that enterprise, I would like to pose two fundamental questions about research in writing assessment. First, what does research tell us about the construction and operation of writing assessments? And second, what are the issues researchers ought to be investigating? In attempting to answer these questions, I want to construe the term "research" broadly to suggest all purposeful, direct efforts to discover some truths about writing assessment, whether those efforts entail hypothesis testing, conceptual model building, the synthesizing of current knowledge, test development, case studies, or other sorts of systematic data gathering. The answer to the first question requires a look at the current state of research knowledge about writing assessment. I will begin there.

A cumulative knowledge base for writing assessment does not yet exist, but the beginnings of one are present in the scattered and so far rather unsystematic research that has been done. The most extensive summary of writing assessment research to date is Leo Ruth's Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment, an NIE-sponsored study done at the University of California at Berkeley. In it Ruth reviews the sources of knowledge for designing writing assessment prompts--psychometrics, professional wisdom, discourse theory, and empirical research--and

concludes that there are few answers to the kinds of questions researchers are asking about the nature of the effects of writing prompts on performance. That statement, true in 1981, still holds today. But the pace of writing assessment research is quickening, and more information is being made available to increasing numbers of testers, administrators, and teachers. I see three major situational variables in writing assessment: topic variables, writer variables, and procedural variables. Each has been the subject of recent research.

Topic Variables. Common sense tells us that how a writing prompt is worded makes a difference to a writer. "Describe what you saw is different from "Describe what you felt." Yet some recent research calls common sense as a criterion into question. Karen Greenberg and I each conducted studies, both of which included lexical variations in writing prompts, that produced no evidence to support the contention that small changes in the wording of essay examination topics of otherwise similar construction affect writers or the scores given their essays.

Greenberg designed four kinds of questions that varied according to the cognitive and experiential demands they placed on writers. She hypothesized that topics with low cognitive demands (calling for interpretation rather than evaluation, for example) and high experiential demands (calling for highly personal responses) would elicit essays that got higher scores. But she found no statistically significant effects for any of her variables, which included changes in content and structure as well as lexical changes.

My study, done with Barbara Ash, asked if it matters whether a

topic is addressed to a writer ("you") or is couched in impersonal or neutral terms; and if it makes any difference whether a topic's "charge" is cast as a question or an imperative. The evidence we gathered--both holistic scores and their statistical analysis and the results of a personal review of the essays--suggested the answer to both questions was no. We came away feeling that as long as topics do not require special knowledge and are suited to the characteristics of the test takers, neither small wording changes nor subject matter changes had much of an effect on writers, their essays, or the scores their essays received.

That is not to say that subject matter does not make a difference, either in the classroom or on a writing performance test. But how much and what kind of difference are not at all certain. One of my studies in a test-development project in Florida produced no difference in the mean scores of essays written on topics asking for opinions on a public issue or demanding introspective reports on writers' inner lives.

Thomas Hilgers has suggested that the problem of unequal familiarity with the content of writing prompts can be solved by giving writers enough information to overcome any deficiencies they might have. Carefully constructed prompts, however--prompts that match the characteristics of the writers who use them--work to solve this problem from the beginning. More research on the characteristics of groups of test writers--previous training, education, cultural distinctions, and gender, as examples--could help test makers produce topics that are as uniform as possible, interesting, and fair.

James Hoetker and I have developed a technique for generating a

hypothetically unlimited number of syntactically and semantically similar topics. The technique involves the use of "frame sentences" akin to those developed as a pedagogical device by structural linguists. Because this approach represents something of a new departure in creating writing prompts, I will discuss it in some detail later in this paper.

Some research has suggested that differences in syntactic complexity reflect differences in mode of discourse. While it is generally accepted that argument is more complex and thus more difficult to write than narrative, it was not until 1981 that the Canadian researchers Aviva Freedman and Ian Pringle provided experimental research evidence for that difference. They found that 12- and 13-year old students were able to realize the conventional schema for the structure of a story more readily than that for the structure of an argument.

Some research on transformational sentence combining has prompted the notion that writing is generally better if it is structurally more complex. This position has been attacked theoretically by James Moffett, Michael Holzman, and others. A recent study by Lorraine Neilsen and Gene Piche found that the level of vocabulary rather than the level of syntactic complexity accounted for differences in ratings of written passages. But the research findings on the relationship between syntactic complexity and quality remain mixed.

Length has been found to be an important variable in the scoring of writing samples on performance assesments. Studies by David Thomas and Dan Donlan; by Ellen Nold and Sarah Freedman; and one of my own revealed a high correlation between length and quality of writing. A longer piece is likelier to be perceived as a better piece of writing.

Why some people write longer and often better essays on writing assessments is open to question. One possible answer is the effect of the amount of information given writers. In a study of rhetorical specification in essay examination topics, I found that a moderate amount of information in a topic--a short introductory statement followed by a charge--elicited essays that got higher scores on the average than those prompted by short, open-ended topics or long, fully specified topics cast in scenario form. The reason, I believe, is that the moderately specified topics helped writers focus and organize their essays more effectively. That is important on a timed test.

A study by Stephen Witte at the University of Texas showed some interesting effects of rhetorical specification on writing. As audience specification was increased, the quality of writing done by the college students in the study increased as well. High specification of content also led to higher quality. But when purpose was more clearly specified, the quality of writing decreased. The highest rated papers came from writing on topics in which audience specification was high, content was specified, and purpose was left unspecified. These findings--and those in my own study--cut against the grain of much professional wisdom on the subject of rhetorical specification. It may be that professional wisdom is based largely on classroom practice whereas the research information comes from assessment contexts. The conclusion I draw from these contradictory positions is that teaching writing and testing writing are much more divergent enterprises than they appear. These divergences ought to be investigated and accounted for.

The IEA Study. Among the most comprehensive research projects on composition topics is that being conducted by the International Association for the Evaluation of Educational Assessment (IEA) under the direction of Alan Purves at the Curriculum Laboratory of the University of Illinois. Purves and his colleagues have developed a model of the composition assignment that takes the form of a classification system setting forth fifteen dimensions of a writing assignment: instruction, stimulus, cognitive demand, purpose, role, audience, content, discourse specification, tone and style, preparation, length, format, time, number of drafts, and criteria for evaluation. The categories are intended to be descriptive and to give test makers and teachers alike a set of tools for "adjusting" writing topics--making them more or less specific--according to the purposes of them. The system can be of use to researchers as well by suggesting a number of factors that may make a difference in writing performance and that can be carefully investigated. It may also help in establishing to what extent the writing assignments in different research studies are comparable. Purves cautions that the IEA model is a highly tentative heuristic device that should be further tested and developed. But it is nonetheless an excellent starting point for identifying those variables in writing assignments that can reasonably be said to make a difference, for characterizing and illustrating those variables, and for specifying their main values.

Writer Variables. Elements of culture, gender, ethnicity, language, psychology, and experience all bear on how different people respond to a writing task. Research with an anthropological, sociological, linguistic, or psychological cast is potentially valuable to our understanding of writing assessment. Two examples of writer variables are

topic interpretation and writer anxiety.

The sociologist William Labov has called attention to the various ways in which an interviewee may interpret a simple question from an interviewer: as a request for information, a command for action, a threat of punishment, a meaningless utterance, or a request for display, among others. The possibilities for such interpretations are just as likely to occur in a writing assessment. Although the writing stimulus may be the same for every examinee, each writer's understanding of it may be different. In survey research, this phenomenon is referred to as "response error." Leo Ruth and Sandra Murphy have developed what they call a constructivist model for interpreting writing tasks--a model of the writing assessment episode beginning with the designing of the writing task by the test maker through the response to the task by the test taker to the evaluation of the writing product by the test rater. From their work, Ruth and Murphy derive four generalizations about the interpretation of writing tasks:

1) Different writers come away with different notions about the significance of particular features of a topic.
2) There are differences in recognition and interpretation of salient points in a writing topic on the part of writers and evaluators.
3) Writers may construct different writing tasks for themselves at different stages in their development.
4) Differences in writers' responses to a writing prompt require further investigation and new approaches to the field testing of assessment instruments. (Such as interviewing writers about how they interpreted a topic.)

While we still do not know very much about how writers read and interpret writing prompts, we do know it is foolish to expect all writers to read and interpret a writing prompt the same way.

The literature on test anxiety is large, but much of it is not relevant to writing assessment because researchers have not investigated apprehension across a variety of writing types and situations. In fact, previous research has drawn measures of writing performance solely from narrative and descriptive essays. One study by Lester Faigley, John Daly, and Stephen Witte found that highly apprehensive writers scored lower on most objective measures of writing-related skills, but lower on only one of two essay tests. There were statistically significant effects of anxiety on personal narrative essays but none on argumentative essays. The researchers speculate that arguing objectively for a particular point of view may create less anxiety than expressing personal feelings--an interesting finding in light of the fact that writing argument is generally considered a more complex task than writing personal narrative.

Procedural Variables. Writing assessments often share certain similarities of scope, design, and instrumentation, typically including a writing stimulus, instructions, a common environment such as a classroom or a meeting hall, a time limit or guide, and a scoring system in which trained raters score the written products in accordance with some established criteria and procedures. All of these factors can affect the outcomes of writing assessments, and some recent research indicates which procedural variables are likeliest to have certain effects.

Sarah Freedman explored the relative effects of three sources of

variables on essay scores--essay, reader, and environment--and found that the essay itself contributed more to the variance in scores. The next largest influence was the training of essay readers: raters trained by trainers who discussed the topics in detail generally gave higher scores than those whose trainers simply discussed the meaning of a topic. Thus we have some evidence for thinking that it is possible to train raters, inadvertently or purposely, to render higher or lower scores.

Rosemary Hake and John Williams found that deliberately nominalized versions of essays (those written in an indirect, verbose, complex style) elicited higher scores and more favorable comments from experienced high school, junior college, and university writing teachers than did verbal versions of the same essays (written in a simple, direct, concise style). Not only is it distressing to note that the raters in the study had praise for the kind of writing they claim as teachers to condemn, but it is eye-opening and unsettling to note that writing performance tests can apparently be subject to factors lying outside the control of test administration--factors, that is, like the conditioning of raters' outlooks on writing that may not be amenable to training.

Research has addressed the question of standards in writing assessments largely in terms of technical issues, such as the statistical reliability of raters' scores, correlations between direct and indirect measures of writing performance, the desirability of one kind of scoring system over another, the number of points on a rating scale, and, in one case, the relationship between the essay scores given out by university graders and the scores those graders received on a per-

sonality questionnaire. Large-scale writing assessment is supposedly a response to the call for higher standards of writing achievement; the standards upon which writing tests are constructed can and should themselves be the subject of research.

The answer to the second question--what are the issues researchers ought to be investigating?--is largely a matter of one's perspective on writing assessment. One viewpoint sure to be represented in discussions of the subject is that of the psychometrician, who wants to refine the technical procedures for operating writing tests: improving the reliability of test scoring, equating various forms of tests and test scores, ensuring that field-test samples are large enough to validate topics, and so forth. These are not unworthy goals, especially given the possibility of a large-scale writing test having to undergo the scrutiny of a law suit, but they are goals that are almost never in keeping with attempts to understand the cognitive processes in writing assessment. As Leo Ruth has pointed out, psychometrics strives to be scientific and to emphasize the manipulation and treatment of products in measurement schemes. Constructing and validating writing assignments, on the other hand, is a subjective and scientifically crude business, particularly as it attempts to understand--"guess at" might describe it better--the cognitive processes involved in reading and responding to a writing prompt. Psychometrics offers little in the way of empirically determined principles for the formulation of essay test questions. So while statisitical procedures can help determine the difficulty of a writing task, they do not help us understand the properties of writing tasks that affect their interpretation by writers responding to them. Research of this sort--investigations into the

interaction between writers and various writing assignments--constitutes one of the most pressing needs in writing assessment today. We simply do not know much about it.

Another kind of useful research knowledge could come from studying the relationships between the goals of writing assessment and the kinds of measures of writing performance provided by assessments. The outcomes of writing assessment can be used for numerous purposes: achievement, competence, evaluation of growth, diagnosis of strengths and weaknesses, placement and prediction, among others. The measures of writing assessment are numerous as well: holistic scoring in its several versions; analytic scoring; primary trait scoring (as well as secondary and tertiary trait scoring); and a host of methodologies for indicating rhetorical properties of written discourse, such as syntactic fluency and cohesion and coherence, and lower order skills, such as usage, grammar, and mechanics. Research into how accurately particular assessment measures represent states of writing performance would be valuable and could pave the way for development of new, more sensitive indicators of growth, achievement, or competence in composing.

But the most significant role for research in writing assessment to play is that of building conceptual bridges between the outcomes of assessment and the processes whereby writing is taught and learned. Writing assessment is, after all, the result of the demand for measures of achievement, and achievement is linked directly to instructional practice. Writing assessment ought to reflect our best understanding of how writing occurs and how it is best taught. It ought to proceed, that is, from an understanding of writing as a complex process of discovering and conveying meaning, a process that is often recursive and

that involves rhetorical, structural, and mechanical choices governed by purpose and audience rather than by compositional rules and patterns. The effect of writing assessment ought to be to improve the teaching of writing rather than limit it, as is the unfortunate case in too many schools, to reductive skill-brokering or paragraph development.

At the heart of the relationship between the testing of writing and the teaching of writing is the question of how to create the conditions under which good writing is known or is apt to occur. We hear that testing drives curriculum; indeed in somes places teachers have taken to offering their students practice in the sort of writing that will appear on a particular large-scale test. If writing assessments were geared to writing processes rather than to short salvos of writing production, they might very well stimulate writing teachers to pay more attention to the craft of composing--generating ideas, gathering data, drafting, revising, editing, and proofreading. True, assessments modeled after composing processes would be hard put to allow for as elaborated a process as could occur in good writing classes, but they would get the message out that this is what happens in good writing classes and that this is what should happen in all writing classes. In other words, if testing drives curriculum, we ought to work to see that it drives it in a progressive way. Some states already have writing assessment programs that allow for composing processes, and these tests are administered to students as early as the 5th grade--a strategy consistent with research findings, especially those of Donald Graves, that elementary-school children can and do revise what they write. Process-oriented assessments would afford opportunities for conducting research on the composing

processes of writers in test situations, something we know very little about.

Research demonstrates that apparently logical and persuasive ideas about writing assessment are not always empirically verifiable. Research also proposes that writing done in and for school, writing done on assessments, and writing done on other occasions are related yet compellingly distinct activities. Future research must address the question of how we may begin to understand the nature of this complex but enticing proposition.

In the past few years, researchers have begun to study the effects of variations in essay test characteristics on the nature and quality of student writing samples. Several studies of essay topics have begun to produce evidence that variations in the wording and content of a topic may affect what and how well students write. In such a study, James Hoetker and I developed an approach to developing sets of topics that may be called "content-fair" in that they afford the great majority of examinees the chance to write on a subject they know. In addition, topics written according to our procedures seem successful in reducing opportunities for misinterpretation and in enabling students to identify a subject about which they have something to say.

In 1983, we contracted to supply a set of topics for Florida's College-Level Academic Skills Test (CLAST), an examination all Florida college sophomores must pass to be admitted to junior-year standing. The CLAST test specifications require students to

produce a specimen of "academic" writing rather than, for example, personal narrative or pure description.

Our own prior experience and research and our reading of the literature on creating essay exam topics convinced us that our topics should share several characteristics that would, we thought, reduce the opportunity for interpretations too different from those we intended: they should be brief, cast in a common syntactic pattern, and should share semantic elements as well. In addition, the instructions for using them should also be brief, simple, and should describe the real rhetorical occasion rather than propose fictional purposes and audiences.

Our biggest problem in writing such topics was content, subject matter. On the one hand, different subjects demanded different vocabularies, which introduced more variety than we wanted in length and syntax. On the other hand,no matter what content we introduced, it was obvious that some students would be better equipped to write about it than others.

Our breakthrough came when we realized that the population of examinees had in common that they were just completing two years of college courses, and that the solution to the problem of the students' varying familiarity with any subject we proposed was to allow each student to write on whatever subject he or she felt most comfortable with. From there we quickly conceived of a "master topic" on the pattern of one of those frame sentences developed by the structural linguists, with individual topics in a set being produced by substituting different words or phrases in specified blanks in the pattern.

When we came across the old-fashioned formal definition, it proved to be just what we needed to get from the master pattern to a set of real-world writing topics:

A formal definition has these parts:

1. A category name (noun or noun phrase)--e.g., "A flute";
2. A linking verb--"is";
3. A class specification (adjective phrase)--"a keyed woodwind instrument";
4. Differentiating criteria (relative clause)--"which has a side hole over which wind is blown, etc."

We adapted this pattern to present a topic consisting of the class specification (3) and two differentiating criteria (4), leaving it up to the student to choose a specific exemplar of the category (1) that completed the definition. For example, a topic might read: "A novel/ which many students read/ that may affect them significantly." A student then is free to identify a novel about which he or she knows enough to write cogently, and free as well to supply his or her own definition of "significant effects." The sample topic could be varied as follows:

"A book/ that many students read/ that may affect them harmfully."

"A book/ that many student read/ that may affect them beneficially."

"A course/ that many students take/ that may affect them in important ways."

"A habit or belief/ that many students acquire in college/ that affects their lives in important ways."

Substitute another noun phrase or other differentiating criteria and another set of related topics can be produced:

"A common practice/ in American schools/ that inhibits learning."

"A common practice/ in American colleges/ that should be changed."

"A common practice/ of teachers/ that does not achieve what it is intended to achieve."

"A common practice/ in American elections/ that does not serve the interests of the public."

Of course, the existence of this mechanism does not relieve the test maker of the responsibility for exercising judgment about the relative appropriateness of topics that are generated. A topic like the first sample, with the class specification "novel," might be appropriate for an examination of English majors but not for a general population of students, many of whom may not know what a novel is, let alone be capable of discussing one on a timed test.

We submitted a set of 30 essay topics--15 containing abstract subjects and 15 with subjects that were people, places, and things--to a review panel of college composition teachers. Their discussions of the topics resulted in revisions of the wording that moved the topics closer to our stated ideals of conciseness, comparability, and fairness. We then field-tested the revised topics with a sample of 900 Florida community college and university students under conditions that approximated as nearly as possible the conditions of the actual CLAST administration. The results were encouraging: for the first

time in our experience, scores were roughly normally distributed, rather than skewed toward the lower end of the scale, and there were fewer extremely low scores than might have been expected. There was, on the other hand, no evidence that the topics contributed to an increase in scores in the higher ranges. Student responses to a questionnaire we appended to the test were decidedly positive, with most students agreeing that the topics were fair and helpful.

We read the field-test essays with special attention to how they opened--whether writers quickly developed a thesis or tended to flounder about--and found an encouragingly large number of essays that began more purposefully than most essays written for earlier testing programs with which we were involved. Our tentative hypothesis was, and is, that these writers took advantage of the support offered them by the topic format to recast the language of the topic into a direction-finding statement that positively affected the quality of their essays.

Because the topics were so different from others used in previous CLAST administrations, the test administrators scheduled another field test of them in late 1984. More than 2,700 students from a representative sample of Florida college sophomores wrote on 15 of the topics, and 20 raters with previous essay-rating experience read and scored their essays. After all scoring was completed, the raters held a summary discussion in which they tried to reach consensus about the suitability of each topic in the field test. They were specifically urged to support their recommendations by relating details of a topic to features of the essays written to it. This process emphasized the use of the special skills of composition teachers to find connections between two sets of texts, the topics and the essays written in response

to them.

The raters were initially critical of what they termed "troublesome similarities" in certain pairs of topics. That is, topics that were obviously variations on one another were perceived as topics that were not different at all. One of the most instructive findings in the field test was that sometimes two "indistinguishable" topics produced comparable sets of essays and sometimes they did not. Students apparently misinterpreted one version of a topic as calling for a "generic" discussion of a category, and their resulting essays were notably weaker.

There were factors influencing how students read and responded to the topics that could not predictably be controlled by reducing syntactic differences and providing a helpful format--factors involving class specifications that were somehow inappropriate or unhelpful. Misfiring topics of this sort (which we would hope to detect before the fact the next time around) illustrate the irreducible importance of designating a class specification appropriate to both the students who must write essays and the raters who must read them.

Eleven of the 15 topics "survived" the field test, which validated the frame-topic format for creating CLAST writing prompts. A recent letter to us from Dan Kelly of the University of Florida, CLAST's chief reader, summarizes events that transpired after the field test:

> As you know, we used the new topic format for the first time for the Fall, 1985, administration/reading of the CLAST essays. We were more than just curious about the effects the topics would have on essays, and I am sure you too are quite interested.
>
> We read at three sites, Tallahassee, Fort Lauderdale, and Gainesville. Readers at all sites agreed that:

1. Essays were better developed because writers could address topics that they not only felt they knew more about, but also were more interesting for them.

2. Thesis statements and organization were as well managed as they had been when writers responded to topics used under the old format.

3. Readers were very grateful for a larger variety of different responses which reduced the number of readings of repetitive supporting ideas and details over the two-day period. We believe this factor led to improvements in reader attention and reliability.

Statistical analyses indicate that scores for essays written under the new format were not significantly different from those recorded during previous CLAST scoring sessions.

If we continue to have similar results on subsequent readings--and there is no reason to believe there won't be--we may assume that the new topics are a genuine breakthrough in the art of topic writing and holistically scored essays.

The new-model topics, that is to say, have been judged to have some noticeably positive effects and no observable negative effects.

REFERENCES

Brossell, Gordon. "Rhetorical Specification in Essay Examination Topics." College English 45 (1983): 165-74.

_____, and Barbara Hoetker Ash. "An Experiment with the Wording of Essay Topics." College Composition and Communication 35 (1984): 423-25.

Faigley, Lester, John Daly, and Stephen Witte. "The Role of Writing Apprehension in Writing Performance and Competence." Journal of Educational Research 75 (Sept.-Oct. 1981): 16-21.

Freedman, Aviva, and Ian Pringle. "Why Students Can't Write Arguments." Ottawa: Carleton U. Linguistics Dept., 1981.

Freedman, Sarah W. "Influences on Evaluators of Expository Essays: Beyond the Text." Research in the Teaching of English 15 (1981): 245-55.

Greenberg, Karen. The Effects of Variations in Essay Questions on the Writing of CUNY Freshmen. New York: City U of New York Instructional Resource Center, 1981.

Hake, Rosemary, and Joseph Williams. "Style and Its Consequences: Do as I Do, Not as I Say." College English 43 (1981): 443-50.

Hilgers, Thomas. "Experimental Control and the Writing Stimulus: The Problem of Unequal Familiarity with Content." Research in the Teaching of English 16 (1982): 381-90.

Hoetker, James. "Essay Examination Topics and Students' Writing." College Composition and Communication 33 (1982): 377-92.

_____, Gordon Brossell, and Barbara Hoetker Ash. Creating Essay Examination Topics. Tallahassee: Florida State Department of Education, 1981.

Holzman, Michael. "Scientism and Sentence Combining." College Composition and Communication 34 (1983): 73-79.

Labov, William. "The Logic of Nonstandard English." Linguistics and the Teaching of Standard English to Speakers of Other Languages or Dialects. Ed. James E. Alatis. Monograph Series on Language and Linguistics No. 22. Washington, D.C.: Georgetown UP, 1969.

Moffett, James. Teaching the Universe of Discourse. Boston: Houghton Mifflin, 1968.

Neilsen, Lorraine, and Gene Piche. "The Influence of Headed Normal Complexity and Lexical Choice on Teachers' Evaluation of Writing." Research in the Teaching of English 15 (1981): 65-73.

Nold, Ellen, and Sarah Freedman. "An Analysis of Readers' Responses to Essays." Research in the Teaching of English 11 (1977): 165-74.

Purves, A. C., A. Soter, S. Takala, and A. Vahapassi. "Towards a Domain-Referenced System for Classifying Composition Assignments." Research in the Teaching of English 18 (1984): 385-416.

Ruth, Leo. Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. NIE Final Report G-80-0034. Berkeley, California: Bay Area Writing Project, 1982.

____, and Sandra Murphy. "Designing Topics for Writing Assessment: Problems of Meaning." College Composition and Communication 35 (1984): 410-22.

Thomas, David, and Dan Donlan. "Correlations between Holistic and Quantitative Methods of Evaluating Student Writing, Grades 4-12" ERIC Doc. ED 211 976. Washington, D.C.: GPO, 1982.

Witte, Stephen, Paul Meyer, Roger Cherry, and Mary Trachsel. Holistic Evaluation: Issues, Theory, and Practice. New York: Guilford Press, in press.